Labour Economics 46 (2017) xxx–xxx

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

High wage workers and high wage peers☆

Michele Battisti

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich, Germany

ARTICLE INFO

Keywords:
Peer effects
Linked employer-employee dataset
Wage differentials

ABSTRACT

This paper investigates the effect of coworker characteristics on wages, measured by the average person effect of coworkers in a wage regression. The effect of interest is identified from within-firm changes in workforce composition, controlling for person effects, firm effects, and sector-specific time trends. My estimates are based on a linked employer employee dataset for the population of workers and firms of the Italian region of Veneto for years 1982-2001. I find that a 0.1 increase in the average labour market value of coworkers' skills (which is around one within-person standard deviation) is associated with a 3.6 percent wage premium. I also find that a sizeable share of the wage variation previously explained by unobserved individual and firm heterogeneity may be due to variation in coworker skills. An event-type study, a Placebo exercise and a series of heterogeneity analyses lend credibility to the baseline results. I also evaluate the role of the spillover effects for wage differentials between specific groups of workers. I find that around 12 percent of the gender wage gap and 10 to 16 percent of the immigrant wage gap can be explained by differences in coworker characteristics.

1. Introduction

It has long been hypothesized that there may be externality effects among people working together (Marshall, 1890, p. 12). Learning about spillover effects among coworkers is important for our general understanding of how labour markets function. In addition, it sheds light on the results of previous work, such as Abowd et al. (1999), that firms are important determinants of wage variation across workers, after controlling for individual characteristics. This topic is increasingly important as firm segregation by worker characteristics (the extent to which certain firms hire certain kinds of workers) has been rising in many OECD countries over the last few decades (Kremer and Maskin, 1996 and Hellerstein and Neumark, 2008 for the US; Kramarz et al., 1996 for France; Lopes de Melo, 2009 for Brazil; Bagger and Lentz, 2014 for Denmark) and may play a role for the recent growth in wage inequality, as found in Edin et al. (2007).

This paper investigates the presence of spillover effects in wages operating between employees of the same firm. In particular, I estimate a log-linear wage regression that adds spillover effects to the person and firm effects model of Abowd et al. (1999). My regression includes fixed individual effects capturing the return to time-invariant worker characteristics, and fixed firm effects that control for unobserved firm-level heterogeneity. I include spillover effects through a measure of coworker characteristics, parameterised as the average of the fixed individual effect among people working at the same firm in the same time period. This represents a proxy measure for the labour market value of coworkers' "portable" skills (i.e., the returns to characteristics that are person-specific and employer-invariant). I estimate the spillover effect arising from coworkers' observable and unobservable time-invariant characteristics simultaneously with the other parameters, using an estimator based upon Arcidiacono et al. (2012). The spillover effect is identified from changes in the composition of the workforce for the same worker in the same firm, controlling for sector-specific time trends.

I estimate my model using the Veneto Worker History (VWH) dataset, a longitudinal linked employer employee dataset that covers the population of private-sector workers of the Italian administrative region of Veneto for each year between 1982 and 2001, and includes earnings and individual characteristics of all workers inside each firm. I find spillover effects to be an important determinant of wage variation: a 0.1 increase in my measure of coworker 'quality' is associated with a 3.6-percent increase in monthly earnings. This means that increasing coworker 'quality' by one standard deviation is associated with a real earnings gain between two and eight percent (depending on the reference distribution used to construct the standard deviation). I also find that including spillover effects strongly reduces the correlation

☆ Financial support by the Leibniz Association (SAW-2012-ifo-3) is gratefully acknowledged. I would like to thank Simon Woodcock and Brian Krauth for their invaluable help with this project. I would also like to thank the editor Albrecht Glitz and two anonymous referees for their extremely insightful comments and suggestions. Peter Arcidiacono and Josh Kinsler have been extremely generous and helpful sharing material from their own work. A special thanks to Krishna Pendakur for access to his computing resources, to Giuseppe Tattara for making the dataset available to me and to Elisabetta Trevisan for precious assistance. I would also like to thank Jennifer Hunt for helpful comments based on a preliminary version of this work. All remaining errors are mine.

http://dx.doi.org/10.1016/j.labeco.2017.02.002
Received 24 February 2015; Received in revised form 12 February 2017; Accepted 17 February 2017
Available online 20 February 2017
0927-5371/

between worker and firm fixed effects, suggesting that workers with ‘better’ labour market skills tend to also have coworkers with ‘better’ skills. In a robustness check, I show that effects are stronger for blue collar workers, which suggests peer pressure as a possibly important mechanism. In an event-type analysis, I track the evolution of wages for workers employed at firms that experience a sudden change in peer quality. The timing of the effects suggests that confounders (firm-specific unobserved trends) are unlikely to introduce sizable omitted variable bias.

I also carry out a placebo exercise, where I find that future peer quality has no effect on current wages once current peer quality is controlled for. This lends credibility to the baseline estimates, suggesting that the main findings are not simply driven by unobserved firm-level time trends. The last part of the paper investigates the role of skill segregation on wage inequality for specific groups of workers in the presence of spillover effects. Based on the empirical setup of this paper, around 12 percent of the gender wage gap and from 10 to 16 percent of the immigrant wage gap may be due to labour market characteristics of peers.

The issue of peer effects in the workplace has recently attracted some interest among empirical economists. However, most of the existing research is based on small datasets on narrow economic sectors and few firms, and focuses on the effect of peers operating through effort and on the role of team production in specific firms. Using panel data from twenty steel mills, Boning and Ichniowski (2007) investigate the effects of the adoption of problem-solving teams, and find a significant positive effect on productivity. More recently, Chan et al. (2012) focus on a different question and investigate the role of compensation schemes on peer effects and on the level of cooperation inside the firm, using data from a Chinese department store. Hamilton et al. (2003) investigate the effect of group composition on the productivity of teams using data from a garment plant, and find evidence of large and heterogenous spillover effects. Bandiera et al. (2009) focus on the effects of social connections between workers and managers on productivity, using data from a soft fruit picking farm, and find that social connections increase the productivity of workers. Ichino and Maggi (2000) look at the role of social interactions on shirking behavior in a large Italian bank, and find group interaction effects to be sizable. On the other hand, Guryan et al. (2009) test for the presence of peer effects in productivity using a dataset of professional golf players, and find no evidence of significant peer effects in that context. Most recently, Arcidiacono et al. (2017) find positive productivity spillovers between teammates in basketball.

Together with Cornelissen et al. (2017), which I further discuss below, to the best of my knowledge this is the first paper that investigates spillover effects in wages using a large dataset that is representative of the overall labour market. Studies on specific firms or occupations tend to find that peer pressure and team-based work matter: observed effort levels are higher when a worker is paired with higher-productivity individuals. The reason for the scarcity of results on the labour market as a whole is related to the complexity of statistically identifying spillover effects, which generates steep data requirements. First, workers in the same firm tend to have similar wages even in the absence of social interactions, simply because they share similar characteristics and because they operate in the same environment. This can generate an upward bias in our peer effects estimates. Therefore, spillover effects ought to be identified from *changes* in workforce composition within firms, for which we need a panel dataset. In addition, some of the relevant coworker characteristics may be unobserved to the econometrician, and their exclusion might result in underestimating the role of spillovers in the labour market.

Until recently, virtually all observational data on the labour market were individual surveys, household surveys or population censuses, making it impossible to link firm characteristics and characteristics of coworkers to any specific worker. Recently, the availability of linked employer employee panel datasets (LEED), which allow researchers to group coworkers together, and to follow the same workers over time, allows to investigate peer interactions within firms accounting for the role of unobservables. Several papers focus on the role of labour market networks on the diffusion of information across workers and labour market outcomes. Recent examples from this literature include Dustmann et al. (2016), Glitz (2017) and Cingano and Rosolia (2012). Battu et al. (2003) measure spillover effects in the UK operating through the level of education of coworkers, but cannot control for the role of unobservables at the worker or firm level. In a related contribution, Navon (2010) investigates the effect of knowledge diversity on within-plant human capital spillovers using data from Israel. Shvydko (2007) specifies the peer effect via coworkers’ wages, which raises endogeneity concerns, since all of the unexplained within-firm wage variation that is common across coworkers affects the estimated spillover coefficient. Lengermann (2002) estimates spillover effects operating through coworker characteristics, similarly to this paper. He finds that a one standard deviation increase in an index of coworker skill is associated with wage increases of three to five percent. Lengermann (2002) uses a different estimator from that of this paper. Its statistical properties are unknown. In particular, Lengermann (2002) cannot show that his estimates are consistent. Cornelissen et al. (2017) is most closely related to this work. They also adapt the estimation strategy of Arcidiacono et al. (2012) to a labour-market application, and estimate spillover effects on wages using average individual fixed effects of a peer group as a measure of coworker quality. Cornelissen et al. (2017) use German social security data. Unlike my data, their data include a detailed variable for individual occupations. They are therefore able to isolate peer effects within very specific groups of workers. Cornelissen et al. (2017) find small peer effects on average (around 2.5 times smaller thank those that I find in this paper, for the most comparable specification using the full sample), but larger peer effects in low-skilled occupations. While I have much more limited information on occupations, I also investigate heterogeneity in spillover effects, and find patterns similar to theirs, i.e. I find spillover effects to be larger for blue-collar and for low-wage workers.

There are some important differences between Cornelissen et al. (2017) and this paper, which can be informative for future work in this field. The region of my data is very different from that of Cornelissen et al. (2017), who use data from Munich. My paper focuses on a region of Italy characterised by relatively low geographic mobility, a prevalence of manufacturing, and by small, specialised firms concentrated in relatively few sectors. Indeed, the industrial districts of Veneto have been studied extensively in the economic literature as an example of the importance of the locally-generated know-how as a source of innovation (Piore, 2009). Whereas Cornelissen et al. (2017) use daily wages as their main outcome variable, I use monthly full-time-equivalent earnings. This means that the spillover effects I estimate are going to be affected by different margins, and may be in part affected by peer pressure that operates through differences in days worked.[1] In addition, the VWH dataset does not have any censoring at the top of the earnings distribution, which might matter, especially for specifications that focus on the upper part of the wage distribution. Differences in the worker and firm population, lack of censoring, differences in the main dependent variable of interest may explain quantitative differences between the results of Cornelissen et al. (2017) and those of this paper. Future work might help our understanding of the relative importance of each of these differences.

2. Background

The theoretical literature has identified a number of channels

[1] Data from the Italian Labour Force survey of 1993-2001 show that variation in days is an important margin of adjustment, which may be affected by peer pressure.

through which the labour market quality of coworkers may affect a worker's wage. First, there may be complementarity effects in the production function, such that a worker's marginal productivity may depend on the characteristics of her coworkers. One channel that has received some attention is the possible effect of human capital heterogeneity at the firm level on productivity and wages, as in Kremer (1993); Davis and Haltiwanger (1991); Kremer and Maskin (1996); Dunne et al. (2000). Navon (2010) finds that knowledge heterogeneity within a firm matters for spillover effects across workers. In a related contribution, Moretti (2004) tests for the existence of human capital spillovers across firms within cities and finds productivity spillovers to be positive and significant for high-tech plants in the United States. Production complementarities, by which the productivity of high-ability workers may affect that of lower-ability workers, may be the result of structural interactions in the production function, as discussed in Guryan et al. (2009) and Moretti (2004). In the empirical analysis below, I perform a few heterogeneity analyses to investigate whether the spillover effects I find are simply the result of production complementarities. Results do not seem to support this claim.

The characteristics of peers might play a role in wage determination in the absence of complementarity effects in the production function. Several articles examine the role of peer pressure in the workplace using laboratory and field data for isolated tasks. Falk and Ichino (2006) use a laboratory experiment to investigate social pressure spillovers, and find that productivity is higher and less dispersed when subjects work in pairs. Mas and Moretti (2009) use field data from a large supermarket chain where worker pairs are varied. Their estimates show that individual effort is positively correlated with the productivity of nearby workers.

The labour market quality of coworkers might also affect individual wages through reservation wages, preferences or social norms. Workers may have a preference for working with certain types of coworkers, and may be willing to accept a lower wage for that because of compensating differentials, and this may generate positive or negative spillover effects. Kremer and Maskin (1996) discuss the possible effects of social pressure on wage equality within the firm. Reference points may also be important for wage determination (Dittrich et al., 2011). If the wage structure within the firm provides a reference point for all workers, wages may be affected by the skill composition of the whole workforce of a firm. For instance, Kronenberg and Kronenberg (2011) find that workers are more likely to leave a firm as wage inequality in the firm increases.

In addition, coworkers' skills may affect wages through bargaining externalities. If high-skill workers are able to extract a higher share of the surplus and bargaining outcomes are positively correlated within a firm, a worker's wage will increase with coworker skills. Conversely, in a context where wages are a fixed share of total revenues, there may be negative bargaining externalities if some groups have a higher bargaining power than others.[2] The expected level of cooperation among workers (and thus total output and individual wages) may also depend on the distribution of types. Investigating spillover effects in wages empirically allows us to assess the relative importance of some of these different channels.

3. Empirical model

My empirical model builds upon the basic structure of the model of Abowd et al. (1999). In the following, let i denote a worker (I sometimes refer to worker i as the focal worker), j denote a firm and t a time period, which is one year in my case. Since the estimation follows workers over time, a more precise notation defines the firm where worker i is employed at time t as $J(i, t)$. I use j for simplicity. A worker i working at a firm j in period t shares that same employer j with other workers, which I refer to as i's set of current coworkers, or current peer group. I denote the set of workers employed by firm j at time t with N_{ijt}, with cardinality N_{ijt}. One of worker i's coworkers is denoted by p. My main regression model is

$$w_{ijt} = \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta + \theta_i + \left(\frac{1}{N_{ijt\sim i}}\sum_{p\in N_{ijt\sim i}}\theta_p\right)\eta + \psi_j + \tau_t + \epsilon_{ijt} \tag{1}$$

where the outcome of interest is worker i's log wage w_{ijt}. I denote time-varying individual and firm characteristics with the $b \times 1$ vector $\mathbf{x}_{\mathbf{it}}$, individual time-invariant characteristics by θ_i, whose average among peers[3] is $\frac{1}{N_{ijt\sim i}}\sum_{p\in N_{ijt\sim i}}\theta_p$. Time-invariant firm characteristics are captured by firm fixed effects ψ_j, while industry-specific time trends are controlled for by τ_t. The $b \times 1$ column vector $\boldsymbol{\beta}$ (where b is the number of individual time-variant characteristics included in the model) and the scalar $\boldsymbol{\eta}$ are parameters to be estimated. The scalar $\boldsymbol{\eta}$ captures the effect of average time-invariant individual characteristics of peers on individual i's log wages, which is the main parameter of interest. Finally, ϵ_{ijt} is a transitory mean-zero error term.

As discussed in Manski (1993) and Bramoulle et al. (2009) there are significant challenges for the identification of peer effects in a linear-in-means model. The steps below are aimed at addressing the main identification challenges. Individual covariates $\mathbf{x}_{\mathbf{it}}$ are included because individual characteristics that have an effect on wages might also be correlated with the average labour market quality of a worker's peer group. I also control for firm size, so that my estimates of peer effects are not driven by changes in the number of employees of a firm that are correlated with average coworker skills. This would be the case if for example firms only managed to attract lower-ability workers when growing quickly. There may also be common-environment effects: some firms might be systematically better at attracting high-wage workers and might also pay higher wages, conditional on a worker's fixed effect. I address this issue by including time-invariant firm effects denoted by ψ_j in Eq. (1).

In addition, there may be trends in the average ability of peers that are correlated with the dependent variable, thereby affecting estimates of spillover effects. For example, during a macroeconomic expansion firms may pay higher wages but may also see the average ability of their workforce decrease, which would be the case if marginal workers had lower-than-average skills. Time trends may be heterogeneous across different segments of the labour market, for example if different economic sectors follow different business cycles, are exposed to different regulatory environment, or are heterogeneously exposed to global competition. In order to control for this, I include industry-specific year effects, denoted by τ_t in Eq. (1). The individual fixed effects θ_i measure the 'market value of portable skills' or 'portable component of individual wages'. For convenience I define $\bar{\theta}_{ijt} \equiv \frac{1}{N_{ijt\sim i}}\sum_{p\in N_{ijt\sim i}}\theta_p$, which is the mean of θ among people working

[2] Incentive schemes within the firm can also generate interactions between wages and peer characteristics. In tournament models, initiated by the seminal work of Lazear and Rosen (1981), effort is a function of the characteristics of all workers in the firm. However, the relationship between labour market quality of coworkers and individual effort may not be monotonic (Becker and Huselid, 1992), because of the *discouragement effect*: low ability workers may choose zero effort if they perceive their probability of winning to be very low. Harbring and Irlenbusch (2003) offer an excellent review of the literature.

[3] Sometimes I refer to this measure as peer 'quality' or 'labour market quality', which I define as a summary measure of time-invariant skills as they are valued by the labour market in terms of wages, similarly to Borjas (1987). The reader should be cautions with its interpretation, however. The parameter θ will capture all of the characteristics that make a worker more productive and the return to those characteristics as well as the characteristics that will make him/her more able to extract rents. My estimates of θ capture the market value of portable skills, and so it does not address the underlying mechanisms through which that market value may be different for different workers.

with worker i at time t, excluding worker i herself. The nonlinear least squares problem derived from Eq. (1) can be written as

$$\min_{\beta,\theta,\eta,\psi,\tau} \sum_i \sum_t \left[w_{ijt} - \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta - \theta_i - \overline{\theta}_{ijt}\eta - \psi_j - \tau_t \right]^2 \tag{2}$$

Eq. (2) is written under a ‘proportionality’ assumption on the characteristics included in θ_i, which is also made in Arcidiacono et al. (2012) and Altonji et al. (2015). This assumption gives a structure to the relationship between the coefficients on each of the components of θ_i in the direct effect on w_{ijt} as opposed to its indirect effect through peers. The proportionality assumption states that the relevant importance of each of these components is the same in the direct effect on own wages and in the peer effect. For example, if two characteristics that are part of θ_i have the same effect on the log wage of worker i, those same two characteristics will also have the same effect when operating through peers.

Under the proportionality assumption, Theorem 1 of Arcidiacono et al. (2012) guarantees consistency and asymptotic normality of $\hat{\eta}_{NLS}$, the nonlinear least squares estimate of η. The key assumption of Theorem 1 requires residuals across any two observations to be uncorrelated (written as mean-independence here for simplicity): $E(\epsilon_{ijt}|\mathbf{x}_{\mathbf{it}}, \theta_i, \overline{\theta}_{-ijt}, \psi_j, \tau_t) = 0$. Net of person effects, firm effects, time effects and spillover effects, the remaining wage variation is assumed to come from transitory shocks. This assumption implies that workers may be systematically different in their unobserved ability, firms may be systematically different in the average ability of their workforce, there might be yearly time trends that are different for different sectors. The remaining intertemporal changes in peer ‘quality’ within a firm, controlling for all of the other covariates, are assumed to be orthogonal to the error term ϵ_{ijt}. This is equivalent to assuming that there are no time-varying unobservables driving changes in the composition of the peer group of worker i while at the same time systematically affecting worker i's wage. One threat to identification may come from unobserved firm-level shocks that are not captured by sector by year trends that affect both individual wages and systematically correlate with changes in the quality of the workforce. In the absence of productivity data at the firm level (for example, data on profits, investment, capital stock etc.), I am not able to directly control for these possible confounders. However, the event-type study and placebo exercises discussed below greatly limit the scope for such concerns.

As discussed in Arcidiacono et al. (2012), under the assumption stated above, the nonlinear least squares solution $\hat{\eta}_{NLS}$ is a consistent and asymptotically normal estimator of the true parameter η as the number of individuals goes to infinity for a fixed number of time periods, even when the underlying fixed effects are not consistent. The key elements that allow Arcidiacono et al. (2012) to prove their theorem is that the vector of individual fixed effects can be written as a function of the spillover parameter and of the data, so that the Least Squares problem above can be formulated as an optimization problem with only one minimand, η. Arcidiacono et al. (2012) can then use Theorem 12.2 of Wooldridge (2002) for consistency of M-estimators establishing identification and uniform convergence, and Theorem 12.3 for asymptotic normality. Even though my setup includes additional sets of fixed effects, the logic of their proofs directly applies.

There are reasons why Eq. (1) is still restrictive. First, the model is specified as a linear-in-means model. This is by far the most common choice in the peer effects literature, with a notable exception being (Brock and Durlauf, 2001). This assumption implies that I cannot investigate spillover effects operating through a different moment of the relevant distribution. In addition, I assume away endogenous effects: peers’ wages affect a worker's wage only through the effect of peers’ ability, not directly via own wages. Without this assumption, my estimates can be viewed as a combination of exogenous and endogenous effects, i.e. effects operating through peer characteristics and behavior. If peers’ effort choice positively affected an individual's effort, and effort and ability were correlated, my estimates of η in Eq. (2) would be biased upwards.

In order to estimate Eq. (2) I find the vector of parameters θ and the parameter η that minimize Eq. (2) iteratively. Estimating Eq. (2) in one step is not computationally feasible with a large dataset. Because of spillover effects, the outcome of person i at time t is a function of the ability of all of i's coworkers, which are themselves estimated within the model (when the θs are updated, all of the other fixed effects and covariates are treated as columns of data. For additional details see Appendix A). I start from a model without spillover effects by setting $\hat{\eta}^0 = 0$ to get a first set of estimates of all fixed effects. I then use these first estimates to get a first set of estimates of the regression parameters β and η. Next, I update the fixed effects to be used in the following step, and proceed by alternatively updating fixed effects and parameters at each step, until convergence is reached. The specific iterative procedure I use builds upon that of Arcidiacono et al. (2012).[4] Each iteration consists of four steps. For a general iteration q, I first estimate $\hat{\eta}^q_{OLS}$ and $\hat{\beta}^q_{OLS}$ from $\boldsymbol{\theta}^{q-1}$, $\boldsymbol{\psi}^{q-1}$, $\boldsymbol{\tau}^{q-1}$ using Ordinary Least Squares. Secondly, I estimate $\boldsymbol{\theta}^q$ from $\boldsymbol{\theta}^{q-1}$, $\boldsymbol{\psi}^{q-1}$, $\hat{\eta}^q_{OLS}$ and $\hat{\beta}^q_{OLS}$ using Eq. (A.2). I then estimate $\boldsymbol{\psi}^q$ from $\boldsymbol{\theta}^q$, $\boldsymbol{\tau}^{q-1}$, $\hat{\eta}^q_{OLS}$ and $\hat{\beta}^q_{OLS}$ using Eq. (A.3), and finally I estimate $\boldsymbol{\tau}^q$ from $\boldsymbol{\theta}^q$, $\boldsymbol{\psi}^q$, $\hat{\eta}^q_{OLS}$ and $\hat{\beta}^q_{OLS}$ using Eq. (A.4).

4. Data

The empirical analysis below uses the Veneto Worker History (VWH) dataset for the years 1982-2001. The region of Veneto is the third largest Italian region by GDP, and the fifth largest by number of residents, with a population of around five million. The VWH dataset has been constructed using the Social Security administrative data of the *Istituto Nazionale per la Previdenza Sociale* (INPS). The dataset includes virtually all private-sector employees of the Italian region of Veneto. It aggregates all establishment identifiers into a firm identifier, which allows me to groups together all workers sharing the same employer. The entire employment history in the period 1982-2001 has been reconstructed for each employee, including employment spell durations and earnings for each spell in each year. Unlike other datasets, there is no censoring at the top of the earnings distribution. This is important in this context, particularly for those part of the heterogeneity analyses where I use individual fixed effects for workers that are in the upper tail of the wage distribution. Additional details on the structure of the VWH dataset and the definition of what constitutes a ‘firm’ in the VWH dataset are available in Appendix B.

Estimating the effects of coworker characteristics on wages requires a certain degree of wage flexibility. Italy is often viewed as a country where collective bargaining is the main mechanism for wage determination. In reality, and especially for small firms, there are many sources of wage heterogeneity across workers. National regulations are typically silent about compensation levels. Trade union contracts specify non-binding minimum wages at the industry level. Although these are relevant for bargaining inside the firm, they only represent an industry-specific floor for total compensation, and in Veneto compensations are almost always higher, as discussed in Bartolucci and Devicienti (2013), who using the same data source find that almost all employees earn a wage premium, and that the median wage premium is around 24 percent. In Italy, individual bargaining is quantitatively important: wage variability within firms is around two thirds of overall wage dispersion (Lazear, 2008). Wage premia are highly heterogeneous across firms (Erickson and Ichino, 1994), and higher for small firms (Cingano, 2003).

In order to estimate my model, it is necessary to identify a specific

[4] This procedure is also used by Cornelissen et al. (2017). Lengermann (2002) includes estimated fixed effects of coworkers as a measure of peer quality, but its procedure ignores the feedback effect of the existence of spillovers, which implies that spillover effects cannot be proven to be consistent.

Table 1
Descriptive statistics from the regression sample: 1982-2001.

	Sample	(1)	(2)	(3)	(4)
		All years	1982	1991	2001
Share of Female Workers		31.75	24.01	29.31	32.60
Share of White Collar Workers		25.11	26.92	28.64	26.17
Share of Foreign Born Workers		7.02	2.42	4.51	10.59
Firm size	p25	19	22	19	19
	p50	58	79	54	55
	p75	319	442	310	243
Gross Monthly Earnings (FTE)	Average	2654	100	120.27	122.64
	p25	2174	100	113.65	114.94
	p50	2733	100	114.80	116.30
	p75	3488	100	121.75	124.84
S.d. log Monthly Earnings		0.570	0.531	0.562	0.579
Share of workers by number of employers in the sample	One	41.56			
	Two	23.60			
	Three	15.12			
	Four	8.96			
	Five or more	10.76			
Number of observations		28,115,529	1,306,253	1,464,793	1,536,351
Number of workers		3,180,714	1,306,253	1,464,793	1,536,351
Number of firms		231,195	60,431	79,176	83,173

Notes: ‘All years’: 1982-2001. Firm size is obtained using a dataset with one observation for each individual, so that it is not representative of the firm distribution. Gross monthly earnings are in 2003 Euros for the full sample, and then are set to 100 for 1982, with the values for 1991 and 2001 being constructed relative to 1982. ‘FTE' above denotes ‘Full Time Equivalent’.
Source: Veneto Worker History Dataset.

time dimension for the panel dataset such that in each time period there is at most one observation for each worker. I therefore construct a dataset where there is at most one observation for each worker in each year, which follows common practice in the literature (Abowd et al., 1999; Card et al., 2013 and Cornelissen et al., 2017 also have one observation per worker per year). I therefore link employment relationships that are recorded over multiple spells but are actually a single spell, and identify the main spell for each worker in each year. Please see Appendix B.2 for additional details.

My main dependent variable is a variable measuring average monthly earnings for full time employment, which is primarily driven by variation in compensation per unit of time rather than by labour supply variations. I construct it using information of total annual compensation, fulltime/part time status and number of months worked. Different papers within the literature use different earning and wage measures, with Abowd et al. (1999) using total annual compensation and other such as Card et al. (2013) and Cornelissen et al. (2017) using daily wages. These choices are typically driven by data limitations. Focusing on monthly earnings rather than total annual compensation implies that differences across individuals are not driven by differences along the extensive margin within the year, e.g. generated by spells that begin or end within the year. However, my outcome variable will be driven by differences in hours and days worked, which allows peer pressure to affect wages through longer hours (this is not something that I can investigate directly in my data since I do not observe hours worked). My measure of coworker labour market skills can be constructed only if the firm has at least two workers. Therefore, I also drop all firms with only one employee, which eliminates around three percent of observations. Separately identifying firm effects and person effects requires employment histories to be sufficiently connected. A brief account on the construction of connected groups is available in Appendix B.3.

5. Results

5.1. Summary statistics of the regression sample

Table 1 presents summary statistics for the regression sample, which has 28,115,529 observations for 231,195 firms and 3,180,714 workers. Across all years, 31.8 percent of workers are female, 7.0 percent are foreign born, 25.1 percent are white collar workers. While there is no clear trend for the share of white collar workers, the shares of females and of foreign-born workers increase between 1982 and 2001. Median firm size (using a dataset with one observation per worker per year) slightly decreases through the sample period. The median worker tends to work for a firm that has around fifty employees.

The main outcome variable that I use are real monthly gross earning (I often refer to them simply as ‘wages’). They are around 2,650 Euros on average (using year-2003 Euros), and increase steadily between 1982 and 1991 (around 20 percent in total for the ten-year period), while they increase only marginally between 1991 and 2001 (less than three percent in total for the period). Median wage growth is 14 percent between 1982 and 1991, and around 1.3 percent between

Table 2
Main regression results.

Dependent variable: individual monthly earnings (FTE), in logs: $ln(w_{ijt})$			
	Models		
Variables	(1)	(2)	(3)
Estimated coefficients of covariates			
Experience		0.013***	0.018***
		(0.000)	(0.000)
Experience^2		−0.001***	−0.001***
		(0.000)	(0.000)
Firm size/1,000		0.013***	0.013***
		(0.000)	(0.000)
Coworker 'Quality' $\overline{\theta}$			0.358***
			(0.002)
Fixed effects			
Standard deviation of the person effect: σ_θ	0.383	0.413	0.389
Standard deviation of the firm effect: σ_ψ	0.230	0.215	0.205
Standard deviation of the time effect: σ_τ	0.170	0.201	0.200
Pseudo R^2	0.716	0.720	0.722
Standard deviations of $\overline{\theta}$			
$\sigma_{\overline{\theta}}$ (overall s.d.)			0.218
$\frac{1}{N}\sum_{i=1}^{N}\sigma_{\overline{\theta},i}$ (average of within-person s.d.)			0.104
$\frac{1}{NT}\sum_{m=1}^{NJ}\sigma_{\overline{\theta},m}$ (average of within-firm s.d.)			0.090
$\frac{1}{NT}\sum_{s=1}^{NJ}\sigma_{\overline{\theta},s}$ (average of within-spell (s) s.d.)			0.053
Matching: Workers, Firms, Coworkers			
$Corr(\theta, \psi)$	0.154	0.160	0.012
$Corr(\theta, \overline{\theta})$			0.420
$N_{obs} = 28,115,529$, $N_{workers} = 3,180,714$, $N_{firms} = 231,195$			

Notes: Approximate robust standard errors clustered at the firm level in parentheses. Heteroskedasticity-robust standard errors are clustered at the level of the firm to account for serial correlation of the errors within the same firm. For all three models, specifications that include sector by province by year effects instead of sector by year effects generate the same estimates (province-specific business cycle effects seem to be unimportant). Labour market experience (measured in years) had to be imputed for part of the sample. See Appendix B.2 for details. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01. *Source:* Veneto Worker History Dataset, for years 1982-2001.

1991 and 2001. Wage inequality tends to increase within the period of our data: while the 25th percentile of the monthly wage distribution gains around 15 percent, the 75th percentile gains around 24 percent. Correspondingly, the standard deviation of log wages increases from 0.53 in 1982 to 0.58 in 2001. Table 1 also shows that there is sizable worker mobility in my sample: around 58 percent of workers switch firm within the dataset, which is a sizeable share given that this is an unbalanced panel and I observe many workers only for very few years. For 24 percent of all workers I observe two employers, for around 15 percent I observe three employers, for around 9 percent I observe four employers, and for around 11 percent I observe five or more employers.

5.2. Baseline regression results

Table 2 presents my main estimates of Eq. (1). Column 1 estimates a model with a firm fixed effect, a worker fixed effect and a year by industry effect only. Column 2 adds controls for firm size and labour market experience (second-order polynomial). Controlling for firm effects, the effect of firm size and labour market experience on wages are small. Large firms are systematically different from small firms. However, firms do not pay systematically higher wages as they grow. In Column 3, I add the average person effect of peers, $\overline{\theta}$. Its estimated coefficient $\hat{\eta}$ is 0.358: a 0.1 increase in coworker 'quality' is associated with a 3.6-percent wage premium.[5] Cornelissen et al. (2017) tend to find smaller effects: in the most comparable specifications (before they add full occupational effects, which are not available in my dataset) they find effects for equivalent changes in coworker skills of around 1.5 percent. While the objectives of the two papers are similar, there are several differences that may play a role. The use of a different outcome variables is likely to matter at least to some extent. Future work on different datasets will deliver additional evidence, thereby helping us better understand the mechanisms at work, and the extent to which estimates depend for example on the specific context or on the choice of a specific dependent variable.

Table 2 includes additional measures of the standard deviation of my variable of interest, which are useful to better assess the magnitude of the effects I find. Using the overall standard deviation of $\overline{\theta}$, which is 0.218, a one-standard-deviation increase in the average person effect of a worker's peers is associated with a wage gain of 7.8 percent. An alternative reference distribution is the average standard deviation of $\overline{\theta}$ within a person's career, which is 0.104. This might be more meaningful since the overall distribution of peer labour market quality in the population may not be the natural reference for considering the changes in coworker composition that workers in my data actually experience. Using this alternative reference distribution, a one-standard-deviation increase in peer characteristics is associated with a wage gain of 3.7 percent. In this case, the conditional wage effect of having a group of peers that is one standard deviation higher than average is similar to the effect of two years of labour market experience. From the perspective of a worker considering a move to a different firm, the relevant measure might be the standard deviation of $\overline{\theta}$ across firms, which is 0.09 (including weights for firm size). The associated wage premium in this case is 3.2 percent. Finally, using the within-spell standard deviation (which is 0.053), picking up the extent to which coworker quality changes for stayers in the same firm we get a wage effect of 1.9 percent. One can view the different interpretations as lower and upper bounds, with the effect of changes of one within-spell standard deviation giving the more conservative effects. In Appendix C.1, I include a robustness check where I run the baseline regression for separate samples depending on firm size, and discuss the results.

The last two rows of Table 2 present descriptive evidence on the correlation between firm and worker fixed effects, which has been often used as a measure of sorting of workers across firms. Column 2 shows that in a model that does not include spillover effects there is a positive correlation between person and firm fixed effects (often taken as a measure of sorting) equal to 0.16. Comparing Column 2 and Column 3 shows that once spillover effects are included in the model, the correlation between individual and firm fixed effects falls greatly and becomes close to zero. On the other hand, the correlation between the individual fixed effect of the focal worker and those of her coworkers is large at 0.42. This suggests that thinking of sorting only as one-to-one matching between workers and firms may be incomplete and potentially misleading. Column 3 distinguishes between the role of firms (as institutions that affect wages irrespective of the workers employed in them) and that of coworkers. The results of Column 3 show that through the lenses of this empirical specification workers with labour market skills that are more valued tend to bunch together in the same firms. Once this type of sorting is taken into account, the evidence for individuals that earn conditionally higher wages to sort into firms that pay conditionally higher wages is weak. The larger correlation found in Column 2 seems to be entirely driven by the firm being a proxy for coworker quality when that is excluded from our model.

[5] I report heteroskedasticity-robust standard errors clustered at the level of the firm to account for serial correlation of the errors within the same firm. Arcidiacono et al. (2012) gives no guidance on how to calculate the exact standard errors. While these are only approximate standard errors, given their size this is unlikely to affect inference. I have run specifications with different clustering levels. Results are unaffected.

Table 3
Variance decomposition exercise.

	1982–2001 (all years)		1982–1986		1987–1991		1992–1996		1997–2001	
Coworker Quality (η)	0.358		0.334		0.339		0.348		0.411	
	Var	%	Var	%	Var	%	Var	%	Var	%
$ln(w_{ijt})$	0.325	100	0.302	100	0.313	100	0.324	100	0.341	100
Variances:										
Individual (θ)	0.152	46.6	0.164	54.5	0.142	45.5	0.134	41.4	0.144	42.2
Firm (ψ)	0.042	12.9	0.043	14.4	0.040	12.8	0.040	12.5	0.044	13.0
Year by Sector (δ)	0.040	12.3	0.010	3.2	0.009	2.9	0.007	2.0	0.012	3.5
Covariates ($X'\beta$)	0.003	1.0	0.004	1.2	0.004	1.3	0.004	1.1	0.002	0.7
Spillover ($\bar{\theta}\eta$)	0.006	1.9	0.005	1.7	0.005	1.6	0.005	1.5	0.007	1.9
Residuals	0.090	27.8	0.091	30.3	0.086	27.4	0.090	27.8	0.094	27.6
Covariances:										
2 $Cov(\theta,\psi)$	0.002	0.7	−0.008	−2.7	0.000	0.0	0.007	2.3	0.008	2.3
2 $Cov(\psi,\delta)$	−0.001	−0.5	−0.001	−0.4	−0.001	−0.3	−0.001	−0.4	−0.002	−0.5
2 $Cov(\theta,\delta)$	−0.029	−8.8	−0.010	−3.4	−0.001	−0.5	0.003	1.1	0.000	0.1
2 $Cov(\psi, X'\beta)$	0.002	0.5	0.003	0.9	0.003	0.9	0.002	0.5	0.000	0.0
2 $Cov(\theta, X'\beta)$	0.000	0.1	0.007	2.3	0.005	1.7	−0.002	−0.7	−0.010	−2.9
2 $Cov(\delta, X'\beta)$	−0.002	−0.6	−0.001	−0.4	0.000	−0.1	0.000	0.0	0.000	−0.1
2 $Cov(\psi, \bar{\theta}\eta)$	0.001	0.2	−0.003	−0.9	0.000	0.0	0.003	0.8	0.003	0.9
2 $Cov(\theta, \bar{\theta}\eta)$	0.028	8.6	0.026	8.5	0.023	7.4	0.021	6.6	0.024	7.1
2 $Cov(\delta, \bar{\theta}\eta)$	−0.010	−3.2	−0.003	−1.1	−0.001	−0.2	0.001	0.3	0.000	−0.1
2 $Cov(X'\beta, \bar{\theta}\eta)$	0.001	0.4	0.002	0.7	0.002	0.6	0.001	0.3	0.000	−0.1
N	28,115,529		6,402,136		7,006,115		7,268,458		7,438,820	

Notes: The columns under the heading 'Var' report variances; the columns under the heading '%' columns report variance as percentages of the variance of the dependent variable. The notation $ln(w_{ijt})$ denotes the logarithm of monthly earnings, full time equivalent. The regressions for the different time periods are constructed using pre-estimated individual and firm fixed effects.
Source: Veneto Worker History Dataset, 1982-2001.

5.3. Variance decomposition

Next, I investigate the role of spillovers in explaining overall wage variation in order to assess their economic importance. For this exercise, I follow Card et al. (2013) in the decomposition of the total variance of wages into variances of the individual components and covariances:

$$\begin{aligned}Var(w_{ijt}) = {} & Var(\theta_i) + Var(\psi_j) + Var(\bar{\theta}_{ijt}\eta) + Var(\mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta) + Var(\tau_t) \\ & + 2Cov(\theta_i, \psi_j) + 2Cov(\theta_i, \bar{\theta}_{ijt}\eta) + 2Cov(\theta_i), \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta) \\ & + 2Cov(\theta_i, \tau_t) + 2Cov(\psi_j, \bar{\theta}_{ijt}\eta) \\ & + 2Cov(\psi_j, \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta) + 2Cov(\psi_j, \tau_t) + 2Cov(\bar{\theta}_{ijt}\eta, \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta) \\ & + 2Cov((\bar{\theta}_{ijt}\eta, \tau_t) + 2Cov(\mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta, \tau_t) + Var(\epsilon_{ijt})\end{aligned}$$

As pointed out in Card et al. (2013), because of sampling errors the estimated variance of the fixed effects may be positively biased, and correlation between the sampling errors will result in negatively-biased correlations between the individual and the firm effects. This suggests caution in interpreting the results below, and the results on sorting above. Comparisons over time are meaningful to the extent that this bias is constant.

Table 3 presents results of this variance decomposition from the same spillover model as in the baseline regression, for all years and for four separate time periods: 1982-1986, 1986-1991, 1991-1996 and 1996-2001. The first rows of Table 3 show that the marginal effect of coworker quality (equal to 0.358 in the full sample) decreases through the period: when I estimate four different regressions I find it to be 0.334 between 1982 and 1986, and 0.411 between 1997 and 2001. The variance of the dependent variable (log of monthly real wages) also increases from 0.302 in 1982-1986 to 0.341 in 1997-2001. Individual fixed effects have the largest variance, followed by firm fixed effects. The variance of the spillover effect is relatively modest at around 1.9 percent of the variance of the dependent variable, but plays an important role as part of the covariance with the individual fixed effect. The covariance between the spillover effect and the firm effect is

Table 4
Symmetry of spillover effects.

Dependent variable: $\Delta ln(w_{ijt})$	(1)	(2)
Change in Coworker Quality between $t-1$ and t	0.097^{***}	0.079^{***}
	(0.013)	(0.017)
Positive Change indicator		-0.003^{**}
		(0.001)
Change in Coworker Quality between $t-1$ and t × Positive Change Indicator		0.064^{***}
		(0.009)
Constant	0.024^{***}	0.024^{***}
	(0.000)	(0.001)
N	20,170,467	20,170,467

Notes: Dependent variable: change in individual monthly earnings (FTE, in logs) between two consecutive years. Robust standard errors clustered at the firm level in parenthesis. In all regressions, we restrict the sample to two-year stayers, and use pre-estimated fixed effects in order to construct our measure of Coworker Quality, and run changes in monthly wages on changes in Coworker Quality, using an interaction term to allow for this effect to differ between positive and negative within-firm changes in coworker quality between $t-1$ and t. This approach follows that of Mas and Moretti (2009). Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: Veneto Worker History Dataset, for years 1982-2001.

Table 5
Spillover effects for blue collar and white collar workers.

Dependent variable: individual monthly earnings (FTE) in logs: $ln(w_{ijt})$					
	(1)	(2)	(3)	(4)	(5)
Sample:	Full	Blue	Blue	White	White
Experience	0.018***	0.019***	0.018***	0.020***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience^2	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm size/1,000	0.013***	0.014***	0.013***	0.014***	0.012***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\bar{\theta}$ Blue Collar	0.279***	0.380***	0.327***		0.171***
	(0.004)	(0.003)	(0.005)		(0.005)
$\bar{\theta}$ White Collar	0.105***		0.073***	0.245***	0.179***
	(0.002)		(0.002)	(0.003)	(0.003)
N	25,623,163	20,070,096	18,205,358	7,544,801	7,417,805
$\bar{\theta}_B = \bar{\theta}_W$ (p-value)	0.000		0.000		0.211

Notes: Standard errors clustered at the firm level in parenthesis. Column (1) reports the results from a regression where I use the full sample and only distinguish the spillover effect by white collar and blue collar workers. In Column (1), the sample is smaller than in the baseline regression because firms that do not have both blue collar and white collar workers are dropped. Columns (2) and (3) are based on the sample of blue collar workers only, while Columns (4) and (5) are based on a sample of white collar workers only. As in Cornelissen et al. (2017), I use pre-estimated individual fixed effects in these regressions. Note that the differences in sample size between the full sample here and that of the baseline regression, as well as between Columns (2) and (3) and between Columns (4) and (5) depend on the fact that not all firms have both blue collar and white collar workers in a given year. $\bar{\theta}_B = \bar{\theta}_W$ (p-value) reports the results from testing whether the difference between the coefficient for the effect of blue collar (B) and of white collar (W) peers is statistically significant. Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.
Source: Veneto Worker History Dataset, for years 1982-2001.

much smaller but tends to increase over time: while it is negative in 1982-1986, it turns to positive in the last two time periods. Next, I present results from a series of heterogeneity analyses and robustness checks.

5.4. Symmetry of the spillover effects

Spillover effects may be systematically different, depending on whether changes in coworker quality are positive or negative. I investigate this in Table 4, adapting the analysis of Mas and Moretti (2009). In particular, using the sample of individuals working for the same firm at time $t-1$ and t, I look at the effect of changes in coworker 'quality' on changes in log wages. I interact changes in coworker quality ($\Delta\theta_{-ijt}$) with an indicator variable that is equal to one if changes are positive, thereby differentiating the impacts of positive and negative changes. My results are qualitatively similar to those of Mas and Moretti (2009): positive changes in coworker productivity are associated with larger responses compared to effects of negative changes. Cornelissen et al. (2017), on the other hand, find spillover effects to be very similar between positive and negative changes, looking at the most repetitive occupations only for this robustness check. The asymmetries I find are not as stark as in Mas and Moretti (2009), however, where the average effect is entirely driven by positive changes. In my case, effects of positive changes are around twice as large in magnitude compared to effects from negative changes. This seems to suggest that both peer pressure and knowledge spillover effects might be at work. Peer pressure may be a more important mechanism than knowledge spillovers, since the latter would imply the effect of negative changes to be small or absent.

5.5. Heterogeneity analysis: blue collar and white collar workers

The baseline results of this study uncover positive and significant spillover effects, measured as the impact of coworker quality on individual monthly earnings. Heterogeneity analysis can then be useful to evaluate the relative importance of different mechanisms that may be driving the baseline results for spillover effects. Next, I divide the labour force of each firm in each year in two groups, and evaluate how spillover effects differ across these groups, similarly to Mas and Moretti (2009). Unfortunately, my dataset has rather poor occupational information: five categories in total, largely measuring the relative position inside the firm rather than actual occupations. While sophisticated analysis by occupation is not possible, I can use the rough occupational variable at my disposal to construct a variable that differentiates two groups of workers, which I refer to as blue collar and white collar. This distinction generates around 68.8 percent blue collar and 31.2 percent white collar workers. Using pre-estimated individual fixed effects, I then construct a measure of coworker quality for blue collar and white collar workers separately, and substitute the overall measure of coworker quality with these two new measures in my spillover regression.

The results of this exercise are presented in Table 5. Column 1 uses our full sample to distinguish between changes in coworker quality of blue collar workers and of white collar workers. I find effects of changes concerning blue collar workers to be stronger than equivalent effects coming from white collar workers. In Columns 2 and 3, I restrict the sample to blue collar workers only. Results of Column 3 show that blue collar workers are almost exclusively affected by changes in the characteristics of other blue collar workers. While changes in the labour market quality of white collar workers also play a role, the magnitude of coefficients is around one forth compared to those for blue collar workers. Columns 4 and 5 repeat the same exercise for white collar workers only. Effects are smaller overall, and differ less between the two groups. For white collar workers, changes pertaining to other white collar workers are found to matter about as much as those pertaining to blue collar workers: the two coefficients of Column 5 are not significantly different (p-value is 0.211). Effects of experience and firm size are very similar across samples and specifications.

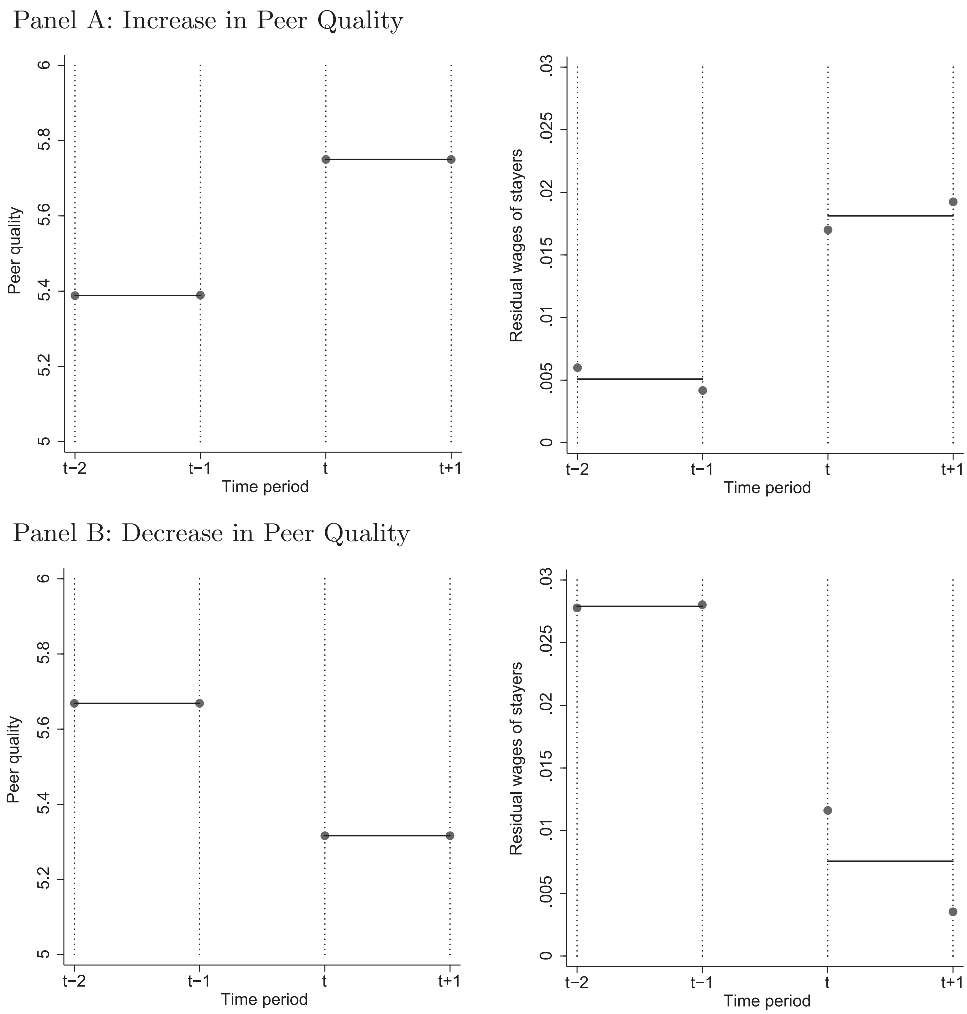

Fig. 1. Peer Quality and Wage Changes. *Notes:* This exercise closely follows that of Cornelissen et al. (2017). The four charts above perform an event-type study to isolate the effect of a sudden change in the peer quality on (residualised) wages. Residualised wages correspond to the residuals of a wage regression that includes controls and fixed effects of our baseline regression, with the exception of peer effects. The charts are constructed from a sample of workers that are employed at the same firm in years t-2, t-1, t and t+1 (i.e. for at least four consecutive years) and for whom peer quality changes very little (less than 0.01 in absolute value) between t-2 and t-1 and between t and t+1, and changes by at least one standard deviation (which is 0.22) in the treatment period, i.e. between t-1 and t. Changes in peer quality over time are defined with the goal of isolating cases in which changes occur in a specific periods and do not exhibit trends. This certainly implies that the results from this exercise are not going to be the same as those of the full sample. The graphs are shown separately for the case of an increase in peer quality (Panel A) and a decrease in peer quality (Panel B). Sample sizes are 1740 for Panel A and 221 for Panel B.
Source: Author's calculations from the Veneto Worker History Dataset.

Overall, while spillover effects are positive and significant in all cases, they seem to be larger for blue collar workers. These results suggest that peer pressure may be important, and that results are unlikely to be simply driven by production complementarities. While results are not directly comparable because of differences in the regression specifications driven by data differences, Mas and Moretti (2009) and Cornelissen et al. (2017) also find larger spillover effects for low-skilled workers and for workers in low-wage occupations. In Appendix C.2, I perform a related exercise. Instead of occupational categories, I generate groups based on the pre-estimated individual fixed effect. I then restrict the regression sample progressively focusing on the top 50, 25 and 10 percent of the within-firm, within-year distribution of individual fixed effects. Results show that spillover effects are not confined to workers with lower labour-market skills, which suggests that production complementarities are unlikely to be the main driver of the baseline results.

5.6. Robustness check: event-type analysis

One of the main concerns with the interpretation of the baseline results is the possible confounding role of unobserved wage trends at the firm level (within each sector), which may be systematically attracting (or repelling) certain types of workers, and may at the same time drive individual wages. In order to address this, I perform an event-type analysis focusing on the evolution of individual wages for workers working in firms that experience a sudden change in peer quality, focusing on workers who stay at that same firm for four consecutive years. Fig. 1 shows results graphically. I look at the effect of a sudden increase and a sudden decrease in peer quality separately. Panel A looks at an increase in peer quality, while Panel B concerns a decrease. These results lend credibility to my baseline specification. Both for increases and for decreases in peer quality, the timing of effects on residualised log wages closely follows the timing of changes in peer quality. In particular, there does not seem to be clearly discernible pre-treatment trends (trends in residualised wages between year $t-2$ and year $t-1$, where the treatment takes place between year $t-1$ and year t). In addition, most of the effect seems to play out at the same time as the change in peer quality, with only small further effects between t and t+1.

The effects of peer quality on (residualised) wages are smaller in magnitude in the event-type study, compared with the baseline results of Table 2. For the same absolute change in 'coworker quality', changes

Table 6
Placebo: current and future coworkers.

Dependent variable: $ln(w_{ijt})$

	(1)	(2)	(3)	(4)
Coworker Quality	Baseline	t+1	t+2	t+3
Current	0.358*** (0.002)	0.355*** (0.005)	0.358*** (0.005)	0.351*** (0.004)
Future ($t+1$)		−0.001 (0.004)		
Future ($t+2$)			−0.006 (0.004)	
Future ($t+3$)				−0.000 (0.003)
N	28,115,529	25,888,913	23,994,412	22,244,136
Current (p-value)		0.427	0.904	0.022

Notes: Dependent variable: individual monthly earnings (FTE), in logs. Standard errors clustered at the firm level in parenthesis. I construct the Peer Quality of Future Coworkers measuring the average of the individual fixed effects of workers employed at time $t+1$ and until $t+3$ at the same firm j where individual i is employed at time t. 'Current (p-value)' reports p-values from comparing the effect of current peers between each of Columns 2, 3 and 4 with the baseline results of Column 1. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: Veneto Worker History Dataset, for years 1982-2001.

in wages are seven to ten times smaller (depending on whether we look at increases of decreases in 'coworker quality') in the event study. The fact that effects are not of the same magnitude is likely to be driven by the fact that the samples of the full regression and that of the event study are quite different. In particular, for the event study to be useful, we need to isolate cases of large sudden changes in 'coworker quality' within the same firm (for the treatment group) and cases where 'coworker quality' remains virtually unchanged over a relatively long period of time (for the control group). This implies that both treatment and control group may not represent all types of firms in the full sample. In addition, workers who stay at the same firm for at least four years (a requirement for my event-type study) are likely to have higher levels of tenure, skills etc, and for some of these reasons also smaller peer effects. Large sudden changes in 'coworker quality' are probably not equally frequent in all types of firms, which also may affect the magnitude of spillover effects we find. Overall, this suggests that while event studies of this type are very useful robustness checks, they are not perfect substitutes for the main analysis on the full sample.

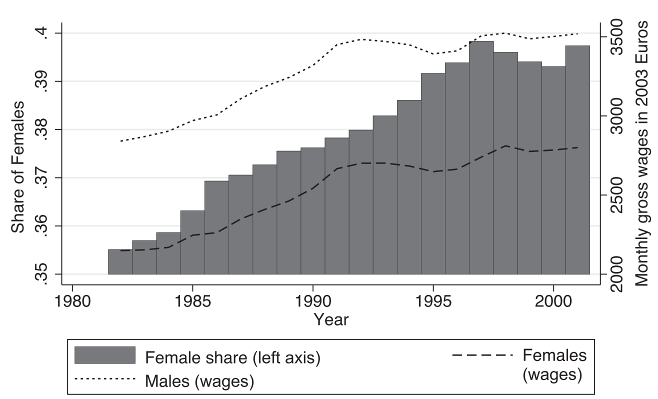

Fig. 2. Average Monthly Wages (full-time equivalent) by Gender and Female Shares. *Notes:* The left axis (and the bars in the chart) reports the share of females in the sample. The right axis (and the lines on the chart) reports full time equivalent gross monthly wages of women and men, expressed in Euros of year 2003.
Source: Veneto Worker History Dataset, 1982-2001, full sample.

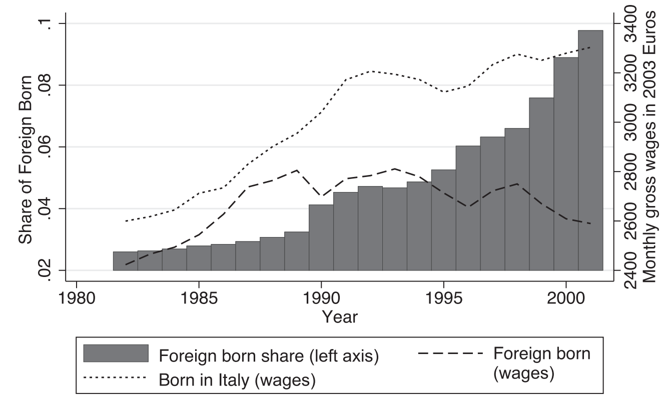

Fig. 3. Average Monthly Wages (full-time equivalent) by Foreign Born Status and Foreign Shares. *Notes:* The left axis (and the bars in the chart) reports the share of foreign born individuals in the sample. The right axis (and the lines on the chart) reports full time equivalent gross monthly wages of workers born in Italy and workers born outside of Italy, expressed in Euros of year 2003.
Source: Veneto Worker History Dataset, 1982-2001, full sample.

5.7. Placebo: current and future coworkers

The event-type analysis above suggests that the spillover effects I find are not simply the result of unobserved trends at the firm level that affect the quality of the workforce and at the same time affect wages of all workers employed at the firm. However, because of the need to isolate cases in which there is a sudden change in peer quality in a specific period while not in the previous or next period, which is a relatively uncommon event, results from the event study may not be representative of the full sample. A simple Placebo-type exercise, similar to that of Cornelissen et al. (2017), may address possible further concerns. In particular, I augment the baseline model of Eq. (1) with a measure of future 'peer quality', parameterised as the average fixed effect of workers employed at the same firm as individual i in periods following period t (excluding worker i if still employed at that firm). Large effects of future peers would be concerning, suggesting that my main specification may suffer from omitted variable bias, i.e. by the presence of unobserved factors affecting the characteristics of hiring and firing, as well as individual wages. The results of this exercise, where I include peer quality for the same firm at times $t+1$, $t+2$ and $t+3$ in different regressions, are presented in Table 6. They do not point to any reason for such concerns: coefficients measuring the effect of future peers on current wages are very small in magnitude and far from statistical significance. In addition, including controls for future peers leaves the estimates of the effects of current coworkers on wages virtually unaffected. Differences in the coefficients are small and statistically significant only for Column 4 (see last row of Table 6),

Table 7
Standardised wage, θ and ψ gaps for different groups.

	log(wage)	Person effect θ	Spillover effect $\bar{\theta}$	Firm effect ψ
Mean	7.88	4.46	4.46	1.78
Standard deviation	0.57	0.39	0.22	0.21
Gender Gap	0.25	0.21	0.08	0.01
Foreign-born Gap	0.13	0.15	0.09	0.02

Notes: 'Gender Gap': difference between average value of men and average value of female workers. 'Foreign-born Gap': difference between average value of native-born and average value of foreign-born workers. Estimates used in this table are the results of the baseline regression, Table 2.
Source: Veneto Worker History Dataset, for years 1982-2001.

which is not surprising given that the sample size of Column 4 is over twenty percent smaller than that of the baseline results.

5.8. Spillover effects and wage differentials

The presence of spillover effects in the labour market suggests that wages are not simply a function of worker and employer characteristics, but also depend on the characteristics of a worker's coworkers. One can then investigate, in a purely descriptive exercise, the extent to which wage differentials of some specific groups of workers are associated with systematic differences in the extent to which these workers have access to higher-'quality' coworkers. Fig. 2 plots average monthly wages (in 2003 Euros, full-time equivalent) by gender, as well as the share of females in my dataset. Monthly real wages increased both for females and males with a break around 1991, with real wages increasing at 2.41 percent a year on average for females and 2.15 percent for males in the years 1982-1991, and only 0.37 percent for females and 0.10 percent for males a year in the period 1992-2001. The gap between monthly wages of males and females decreased slightly from 24.3 percent in 1982 to 20.4 percent in 2001. The proportion of females increases throughout the sample period. Fig. 3 compares workers born in Italy with workers born abroad. The bar chart shows that the proportion of foreign-born workers increases substantially between 1982 and 2001. The unconditional wage gap between foreign born and Italian born was relatively constant in the period 1982-1989. Afterwards, it increases dramatically, driven largely by falling real wages of foreign born. While in 1982-1989 average yearly growth rates of gross real wages are 1.70 percent for Italian born and 1.98 percent for foreign born, in the period 1990-2001 the equivalent figures are 0.71 percent for Italian born and −0.33 percent for foreign born.

Using my baseline estimates, I investigate the extent to which individual, firm, spillover effects systematically differ across groups. Table 7 presents the average of wages and of the estimated fixed effects across genders and immigrant status. On average, female workers have 25 percent lower wages in my data, 20 percent lower 'market value of portable skills' measured by the fixed person effect θ, 8 percent lower coworker 'labour market quality' and work in firms that pay conditionally slightly lower wages. On the other hand, on average foreign born workers have wages that are 13 percent below those of native workers, person effects (θ) are 15 percent lower, coworker 'labour market quality' is 9 percent lower. To the extent that spillover effects affect wages, differences in coworker 'quality' between groups will affect their relative outcomes. Using Eq. (1), the average wage gap between two groups of workers can be written as

$$E(w_{ijt}^{M} - w_{ijt}^{F}) = E[(\mathbf{x}_{\mathbf{it}}^{\mathbf{T}})^{\mathbf{M}}\beta - (\mathbf{x}_{\mathbf{it}}^{\mathbf{T}})^{\mathbf{F}}\beta] + E(\theta_i^M - \theta_i^F) + E(\overline{\theta}_{ijt}^{M}\eta - \overline{\theta}_{ijt}^{F}\eta) + E(\psi_j^M - \psi_j^F) + E(\tau_t^M - \tau_t^F) + E(\epsilon_{ijt}^M - \epsilon_{ijt}^F) \quad (3)$$

where the exponents F and M stand for 'Female' and 'Male' but may refer to any two groups. Based on this decomposition, around 85 percent of the overall wage gap between female and male workers is due to differences in θ, i.e. differences in individual characteristics and their returns in the labour market.[6] Differences in peer 'quality' explain 12 percent of the overall gap (and varies little over time): one eighth of the gender wage gap is due to the fact that females have on average coworkers with lower person effect θ. All other covariates as well as the unexplained component are small. To assess whether differences in the type of coworkers that men and women have depends on their characteristics, I then regress average peer 'quality' on gender and a series of controls:

$$\overline{\theta}_{ijt} = (Female)_i\delta_0 + \theta_i\delta_1 + \mathbf{x}_{\mathbf{ijt}}^{\mathbf{T}}\delta_2 + P_{ijt}\delta_3 + \psi_j\delta_4 + \upsilon_{ijt} \quad (4)$$

where θ, $\overline{\theta}$ and ψ are those I estimated in my main model and *Female* is a dummy for gender. The vector $\mathbf{x}_{\mathbf{ijt}}$ includes a constant, experience and firm size. In addition, P_{ijt} denotes the proportion of females among worker i's coworkers at time t. Finally, υ_{ijt} is a transitory mean-zero error term and δ_0, δ_1, δ_2, δ_3 and δ_4 are parameters to be estimated. Table 8 presents the estimates from Eq. (4). Moving from Column 1 through to Column 5, I gradually include more controls, to evaluate the extent to which the overall differences in average fixed effects across genders are explained by individual and firm characteristics. Column 3 shows that once one controls for the proportion of females among peers, female workers have conditionally higher-θ peers compared to males. Females are not concentrated in peer groups of lower 'quality' once we control for the female share.

Below, I perform equivalent exercises for the immigrant wage gap. Foreign born and native workers are segregated across firms: in 2001, while native workers work in firms where around 9 percent of workers are foreign born on average (the corresponding median is around 5 percent), foreign born workers work in firms where 22 percent of workers are foreign born on average (the corresponding median is 16 percent). A simple decomposition equivalent to that of Eq. (3) shows that the majority of the gap is driven by differences in the person effect θ. Average peer characteristics explains between 10.4 percent in 1982 and 15.9 percent in 1987 of the overall wage gap. My decomposition also shows that a large part of the wage gap (19 percent on average) is explained by the firm effect ψ: foreign born disproportionately work in firms that pay lower wages. I then regress peer characteristics on a dummy for foreign born and on other covariates:

$$\overline{\theta}_{ijt} = (Foreign\ born)_i\delta_0 + \theta_i\delta_1 + \mathbf{x}_{\mathbf{ijt}}^{\mathbf{T}}\delta_2 + P_{ijt}\delta_3 + \psi_j\delta_4 + \upsilon_{ijt} \quad (5)$$

where P_{ijt} denotes the proportion of foreign born among worker i's peer group and all other covariates and parameters are defined as in Eq. (4). Table 9 displays the estimates for Eq. (5). As before, I add additional controls moving from Column 1 to Column 5. Unlike for females, even controlling for own unobserved 'type' θ_i, as well as for the proportion of foreign born among the peer group, experience, firm size and firm effects, on average foreign born still have peers that have lower person effects. Column 5 shows that wages of foreign born workers are around 0.5 percent lower solely due to the characteristics of their peers. Foreign born workers, irrespective of their characteristics, seem to be more likely to work with coworkers less advantageous labour market characteristics. If we believe spillover effects to be an important feature of wage determination, equal access to all segments of the labour market by all workers matters for wage differentials.

6. Concluding remarks

This paper estimates the effect of coworkers' labour market characteristics on wages. As discussed above, I address the main sources of possible bias due to group selection (by which workers with certain characteristics are non randomly distributed across firms) and to the role of unobservables by using within-firm variation in peer group composition net of time trends, and allowing peer effects to operate through all relevant time-invariant worker characteristics. I use a large panel dataset of workers of the Italian region of Veneto for years 1982-2001. Together with Cornelissen et al. (2017), this is the only paper that provides credible estimates of peer effects in earnings for a representative sample of workers. I find peer characteristics to be an important factor for wage determination: a 0.1 point (around one within-person standard deviation) increase in coworker 'quality' is associated with a rise in real monthly wages of 3.6 percent. I check the validity of the main results using an event-type analysis and a Placebo exercise, which suggest that the baseline results are unlikely to be driven by unobserved firm-level wage trends that are correlated with

[6] Note that this component of the gap does not necessarily reflect differences in skills, since it is a combination of skills and their wage returns. It is possible that foreign born workers and female workers have lower labour market skills, but it is likely that they have lower returns to those unobserved labour market skills, for reasons that may include labour market discrimination, as found in a number of audit studies.

Table 8
Gender and labour market quality of peers.

Dependent variable: average person effect of coworkers: $\frac{1}{N_{ijt}} \sum_{p \in N_{ijt}} \theta_p$					
	(1)	(2)	(3)	(4)	(5)
Female dummy	−0.082*** (0.000)	−0.030*** (0.000)	0.037*** (0.000)	0.032*** (0.000)	0.032*** (0.000)
Individual unobserved heterogeneity θ_i		0.247*** (0.000)	0.238*** (0.000)	0.222*** (0.000)	0.221*** (0.000)
Proportion of females in peer group			−0.240*** (0.000)	−0.234*** (0.000)	−0.234*** (0.000)
Experience				0.003*** (0.000)	0.003*** (0.000)
Experience^2				−0.000*** (0.000)	−0.000*** (0.000)
Firm size/1,000				0.026*** (0.000)	0.026*** (0.000)
Firm heterogeneity ψ					−0.017*** (0.001)
Observations	28,115,529	28,115,529	28,115,529	28,115,529	28,115,529
R^2	0.033	0.214	0.285	0.339	0.339

Notes: Fixed effects are from the baseline regression results, see Table 2. Heteroskedasticity-robust standard errors in parentheses. Significance levels: *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Veneto Worker History Dataset, years 1982-2001.

Table 9
Birth place and labour market quality of peers.

Dependent variable: average person effect of coworkers: $\frac{1}{N_{ijt}} \sum_{p \in N_{ijt}} \theta_p$					
	(1)	(2)	(3)	(4)	(5)
Foreign born dummy	−0.094*** (0.000)	−0.056*** (0.000)	−0.014*** (0.000)	−0.014*** (0.000)	−0.014*** (0.000)
Individual unobserved heterogeneity θ		0.254*** (0.000)	0.244*** (0.000)	0.226*** (0.000)	0.226*** (0.000)
Proportion of foreign born in peer group			−0.410*** (0.001)	−0.374*** (0.001)	−0.377*** (0.001)
Experience				0.003*** (0.000)	0.004*** (0.000)
Experience^2				−0.000*** (0.000)	−0.000*** (0.000)
Firm size/1,000				0.025*** (0.000)	0.025*** (0.000)
Firm heterogeneity ψ					−0.023*** (0.001)
Observations	28,115,529	28,115,529	28,115,529	28,115,529	28,115,529
R^2	0.009	0.213	0.240	0.291	0.292

Notes: Fixed effects are from the baseline regression results, see Table 2. Heteroskedasticity-robust standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Veneto Worker History Dataset, years 1982-2001.

my measure of coworker skills. Through the lenses of the my model, one can look at the correlation between the fixed effects of workers inside the same firm as a measure of sorting. It seems that sorting of high-wage workers into better employers operates mostly though the bunching together of employees with similar characteristics.

I also find evidence that spillover effects are not symmetric, with larger magnitudes associated with positive changes in coworker skills. Consistent with the idea that peer pressure may plan a role, additional heterogeneity analysis shows that effects are larger for blue collar workers and for workers with relatively low wages. Overall, we still have a very limited understanding on the mechanisms that are at work behind these results, and how the magnitudes of the effects we find may be affected by differences in the context, types of workers and definition of the outcome variables. In addition, to the best of my knowledge, there is still no empirical evidence on the presence and magnitude of peer effects in productivity that uses data for the labour market as a whole and uses a reliable identification strategy. Therefore, there is a high potential for linked employer-employee datasets that include many more firm-level variables (such as investment, profits, openness measures, internal organisation, financial statements etc.) to provide a fruitful base for future work and provide new insights on important questions, including the effect of worker sorting on firm performance and wage inequality.

Appendix A. Details on the iterative procedure

I define the variable y_{ijt}, which denotes the dependent variable of my model net of all fixed effects and covariates that are not a function of the current θ:

$$y_{ijt} \equiv w_{ijt} - \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta - \psi_j - \tau_t \tag{A.1}$$

As shown below, the key for the estimation is to derive the First Order Conditions of (2) with respect to the worker effect θ_i after having substituted in using Eq. (A.1):

$$\sum_t \left[y_{ijt} - \theta_i - \eta \frac{1}{N_{ijt\sim i}} \left(\sum_{p \in N_{ijt\sim i}} \theta_p \right) \right] + \sum_t \sum_{p \in N_{ijt\sim i}} \eta \frac{1}{N_{ijt\sim i}} \left[y_{ijt} - \theta_p - \left(\eta \frac{1}{N_{ijt\sim p}} \sum_{k \in N_{ijt\sim p}} \theta_k \right) \right] = 0$$

In order to make this implicit equation for θ_i operational, I solve the equation above for θ_i moving all of the terms including θ_i to the left-hand side of the equation and then solving for θ_i:

$$\theta_i = \frac{\sum_t \left[y_{ijt} - \eta \frac{1}{N_{ijt\sim i}} \left(\sum_{p \in N_{ijt\sim i}} \theta_p \right) \right]}{\sum_t \left(1 + \eta^2 \frac{1}{N_{ijt\sim i}} \right)} + \frac{\sum_t \sum_{p \in N_{ijt\sim i}} \eta \frac{1}{N_{ijt\sim i}} \left[y_{ijt} - \theta_p - \left(\eta \frac{1}{N_{ijt\sim i\sim p}} \sum_{k \in N_{ijt\sim i\sim p}} \theta_k \right) \right]}{\sum_t \left(1 + \eta^2 \frac{1}{N_{ijt\sim i}} \right)} \tag{A.2}$$

The person fixed effects that are on the right-hand side of the equation above are those of the previous iteration, and get updated after each θ_i is updated using Eq. (A.2). As a consequence, even though my model includes different and additional fixed effects, Theorem 2 in Arcidiacono et al., 2012 applies here, since the additional estimated coefficients do not depend on theta and thus can be viewed as part of the dependent variable at each iteration. Theorem 2 shows that Eq. (A.2) is a contraction mapping, guaranteeing convergence of the estimated parameters to their NLS counterparts, for any initial vector θ_0 if $\eta < 0.4$.[7] Unlike similar two-step procedures, the presence measurement error in the covariates does not lead to an attenuation bias of the regression coefficients. Arcidiacono et al. (2012) derive this result by stacking the first-order condition from the optimization problems for each θ and checking the conditions for the function from one guess at the vector of individual effects of θ to the next $f\colon \theta \to \theta'$ to be a contraction mapping, which is equivalent to checking the conditions for $\rho(f(\theta), f(\theta') < \beta\rho(\theta, \theta')$ for some $\beta < 1$ and where ρ is a valid distance function. In each step of the iterative procedure, after having updated each member of the vector $\boldsymbol{\theta}$ using (A.2) the procedure updates the firm fixed effect and the year by sector fixed effect averaging the residuals for each observation over the relevant set of observations, excluding the fixed effect of interest. One can then update firm effects and time effects:

$$\psi_j = \frac{\sum_{i \in N_j} \left[w_{ijt} - \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta - \theta_i - \eta \frac{1}{N_{ijt}} \left(\sum_{p \in N_{ijt}} \theta_p \right) - \tau_t \right]}{\sum_{i \in N_j} 1} \tag{A.3}$$

$$\tau_t = \frac{\sum_{i \in N_t} \left[w_{ijt} - \mathbf{x}_{\mathbf{it}}^{\mathbf{T}}\beta - \theta_i - \eta \frac{1}{N_{ijt}} \left(\sum_{p \in N_{ijt}} \theta_p \right) - \psi_j \right]}{\sum_{i \in N_t} 1} \tag{A.4}$$

For updating θ_i I use a modified version of Eq. (A.2) for computational convenience, using the result in Lemma 2 of Theorem 1 of Arcidiacono et al. (2012):

$$\theta_i^q = \frac{\sum_t \left\{ \eta \frac{1}{N_{ijt}} \left(\sum_{j \in N_{ijt}} e_{jt}^{q-1} - e_{it}^{q-1} \right) + e_{it}^{q-1} + \left(1 + \eta^2 \frac{1}{N_{ijt}} \right) \theta_i^{q-1} \right\}}{\sum_t \left(1 + \eta^2 \frac{1}{N_{ijt}} \right)} \tag{A.5}$$

where e_{it} denotes the regression residual from the OLS regression estimates of step 1. Eq. (A.5) is obtained from Eq. (A.2) by identifying regression residuals and then substituting them in, isolating the terms that include θ_i^{q-1}. The residual sum of squares falls by a decreasing amount after each iteration, until a predetermined criterion for convergence is reached (I set the difference in the sum of squared residuals to be less than 10^{-7}

[7] The result in Arcidiacono et al. (2012) is not a bivariate relationship, so that the result may hold for values larger than 0.4 as well, depending on the size of peer groups.

between two consecutive iterations).

Appendix B. The VWH dataset

B.1. Structure of the dataset

The period covered by the VWH dataset is 1976-2001. Because of coding errors for the first few years, I only use the 20-year period between 1982 and 2001. The VWH dataset has not been updated for years after 2001. State and local government employees, farm workers and some category of professionals, such as doctors, lawyers, notaries and journalists, are not included because they have alternative social security funds. Self-employed are also excluded. Additional information on the dataset are available in Card et al. (2014) and in Tattara and Valentini (2010). The firm is identified by a firm tax number, which unfortunately does not allow the identification of separate establishments within the same firm.

The VWH dataset is composed of a worker archive, a firm archive and a job archive. I link the job archive to the worker archive using the worker identifier they share, and the firm archive to the dataset using the firm identifier. The worker archive includes a person identifier, and limited individual information: gender, birth date, birth place and residential address. The VWH dataset includes scarce information on occupations (only a measure of level inside the firms, which is likely to reflect promotions rather than task allocation). Educational attainments are also absent. This is not crucial for my estimation however, because all of the time-invariant individual characteristics are captured by the person effect, and could not be separately included even if available. The firm archive includes a firm identifier, firm's name, activity, address, sector (firms are classified according to the three-digit Ateco 1981 standard classification), establishment date, cessation date, number of initial employees, area code and postal code of the headquarter. The job archive includes a worker identifier, a firm identifier, duration of the employment relationship (in days), place of work, total yearly real wages in 2003 Euros for each job in each time period, qualification, contract level.

For the analysis in this paper, it is crucial to have a correct identification of firms, in a cross sectional as well as dynamic sense. The VWH dataset has been the product of a careful identification of firms as economic entities. The variable has been constructed using the same technique as in Occari and Pitingaro (1997). When more than half of the workers of a large firm moves to another firm the mobility is considered spurious, i.e. the two firms are coded as the same firm. For small firms the logarithm also requires that location remains unchanged (Tattara and Valentini, 2010).

B.2. Construction of the regression sample

From the raw VWH data, I construct a sample with at most one observation for each worker in each year. Apart from cases with missing values in the variables used in the regression, the vast majority of these case are cases in which there are two different records for the same worker in the same firm, which is the result of the fact that the data is based upon a firm identifier that does not take mergers and acquisitions into account. For all cases in which a worker is observed more than once in the same firm in the same year I construct a new relationship that incorporates these different relationships and drops duplicates. For the cases in which there are still multiple observations per worker/year I identify a dominant job keeping the employment relationship with the higher number of days paid.

My main regression model includes a measure of firm size, which I construct counting all employees in a certain firm for each year. This measure may underestimate actual firm size since a firm's workforce may include undocumented workers, or may hire professionals that are not present in the VWH dataset. I also construct a variable for labour market experience: within the period of my data, I can see the employment history of all workers and so I can use the total number of months worked to construct a measure of actual labour market experience. However, for a portion of my sample I cannot observe the full labour market history. For this purpose, I divide workers into two categories, depending on whether I can assume that I observe them from the beginning of their careers. I assume that I see their whole careers if they have no job in the first three years of my dataset and if they are at most 18 years old in 1985. For the workers for whom I assume that I am observing their whole labour market career, experience will be equal to observed experience, given by the sum of months in full time employment up to (not including) year *t*. For workers that I do not see from the start of their careers, experience is given by observed experience up to year *t* plus the average months of experience accumulated by workers of the same category and gender from their average minimum age of employment up to the first time I see them in my dataset. I divide workers into white collar and blue collar workers based on their occupation, in order to control for the different age of entry in the labour force of white collar workers. Each year, male workers work on average around 10 full-time months if they are white collar workers, around 9.5 months if they are blue collar workers. Female workers work around 9 full-time months if they are white collar and around 8.5 months if they are blue collar workers. Average age of entry in the labour force is very similar for male workers and females workers, at around 22 for white collars, 19 for blue collars.

B.3. Connected groups

In order to identify groups of connected observations we need to identify observations that are members of a connected graph. A connected group of firms and workers contains all the workers that ever worked for any of the firms in the group and all the firms where any of the workers were ever employed (Abowd et al., 2002). I adapt an algorithm developed in Ouazad (2007) to identify connected groups of observations. The basic functioning of the algorithm mirrors the definition of connected groups: starting from a single firm, the algorithm finds the set of workers that worked for that firm in any time period, and includes those as part of the connected graph. The algorithm then adds all firms that ever employed our set of workers, and then all workers ever employed by those firms to the connected graph. This procedure continues until no additional worker is added to the connected graph. Abowd et al. (2002) then proceed by estimating person and firm effects within each group to maintain the representativeness of the sample. I drop all observations that are not part of my main connected sample before estimating my model, since only around 9,000 observations out of over 28 million are excluded from the main connected group.

Appendix C. Additional robustness checks

C.1. Small firms and large firms

Table C.1 reports estimates of Eq. (1) run on a sample of workers of small firms, and of very large firms. This can be useful since organizational

Table C.1
Separate regressions by firm size.

Dependent variable: $ln(w_{ijt})$			
Sample	(1) Full	(2) Small firms	(3) Large firms
Experience	0.018	0.023	0.018
Experience^2	−0.001	−0.001	−0.001
Firm size	0.000	0.004	−0.000
Coworker 'Quality' $\bar{\theta}$	0.358	0.184	0.340
Fixed effects			
σ_θ	0.389	0.467	0.443
σ_ψ	0.205	0.372	0.280
σ_τ	0.200	0.171	0.237
Pseudo R^2	0.722	0.813	0.810
Standard deviations of $\bar{\theta}$			
$\sigma_{\bar{\theta}}$ (overall s.d.)	0.218	0.372	0.181
$\frac{1}{N_t J_t}\sum_{j=1}^{J} N_{jt}\sigma_{\bar{\theta},j}$	0.089	0.158	0.056
N_{obs}	28,115,529	3,933,459	4,224,592
$N_{workers}$	3,180,714	1,026,651	683,624
N_{firms}	231,195	203,543	178

Note: The dependent variable denotes individual monthly earnings (full time equivalent), in logs. Small firms are those with less than ten employees; large firms are those with more than one thousand employees. The expression $\frac{1}{N_t J_t}\sum_{j=1}^{J} N_{jt}\sigma_{\bar{\theta},j}$ denotes weighted average of within-firm s.d. of 'peer quality'. For small firms and large firms, my convergence criterion is based on a differences between log likelihoods of successive steps smaller than 10^{-4}; samples are restricted to observations in the main connected group. All effects are significant at the one percent level.
Source: Veneto Worker History Dataset, for years 1982-2001.

Table C.2
Robustness check: high-θ workers.

Dependent variable: monthly earnings (FTE), in logs: $ln(w_{ijt})$				
Sample	(1) Baseline	(2) Top 50	(3) Top 25	(4) Top 10
Experience	0.018***	0.019***	0.022***	0.025***
	(0.000)	(0.000)	(0.000)	(0.000)
Experience^2	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm size / 1000	0.013***	0.014***	0.015***	0.015***
	(0.000)	(0.000)	(0.001)	(0.001)
Coworker 'Quality'	0.358***	0.273***	0.201***	0.143***
	(0.002)	(0.002)	(0.003)	(0.003)
N	28115529	15235937	7347145	2663471
Standard deviation (s.d.) $\bar{\theta}$	0.218	0.261	0.289	0.369
Effect of one s.d. change in $\bar{\theta}$	7.8%	7.1%	5.8%	5.3%

Notes: Standard errors clustered at the firm level. Column (2) reports the results from a regression that includes, for each firm in each year, the top 50 percent of individuals after sorting them by their worker fixed effect θ. In Columns (3) and (4), the sample is restricted to the top 25 and 10 percent workers (in terms of their individual fixed effect) respectively. Coworkers only include individuals that are part of the respective samples. Differences in the effect of 'Coworker Quality' between regressions are significant at all conventional significance levels. Sample sizes fall more than proportionally because of the restriction that we need at least two employees for each firm. Significance levels: * p < 0.1, ** p < 0.05, *** p <0.01.
Source: Veneto Worker History Dataset, for years 1982-2001.

structure and social interactions are likely to be very different between small and large firms. Both in the estimates for small and for large firms I find smaller coefficients for peer effects and smaller partial R-squared. This suggests that my main estimates are not driven by very small firms or very large firms alone. The second column of Table C.1 shows estimates obtained using the sample of firms with less than ten employees. For this sub-population, a one standard deviation increase in the average labour market skills of peers is associated with a wage gain of 6.8 percent. Using the average within-firm standard deviation, the equivalent gain is 2.9 percent. The third column of Table C.1 shows estimates for a sample of the largest firms only. Compared to the full sample, peer effects are smaller: while a unitary change in the overall standard deviation is associated with a wage increase of 6.2 percent, the estimate using average firm-level standard deviation in "peer quality" is of 1.9 percent. The fact that I find smaller effects for larger firms is comforting, since for larger firms the entire set of coworker represents a noisier proxy for the group with whom the focal worker actually interacts.

C.2. Heterogeneity analysis: high-θ workers

The baseline results may be driven by production complementarities, which generate effects of coworker characteristics on wages due simply to structural reasons without any actual interaction among workers. These issues are discussed at length in Guryan et al. (2009) and Moretti (2004).

While it is not possible to directly test for production complementarities in my data, heterogeneity analysis may shed light on their role. Production complementarities are likely to be stronger for workers with lower labour-market skills and weaker for managers and white collar workers. I can then run our spillover effects model for the sub-sample of high-skilled workers. Table C.2 presents the results of such exercise, where sample restrictions are based upon individual fixed effects. In all cases, coworkers are also restricted to be part of the selected sample. Column 1 presents our baseline results again for comparison. In Column 2 we restrict the sample to the employees in the top half of the within-firm, within-year distribution of individual fixed effects. Spillover effects are smaller, with the effect of one standard deviation change (I am using only the overall standard deviation here) falling from 7.8 percent to 7.1 percent. In Column 3 we restrict the sample further to the top 25 percent high-θ employees. The estimated spillover effects fall further, consistent with the view that peer effects are stronger at lower levels of the skill distribution. The effect is however still substantial, with a one-standard-deviation change resulting in a wage effect of 5.8 percent. Column 4 uses the sample of the top ten percent employees in each firm in each year. Again, spillover effects are somewhat smaller but still sizeable. Overall, while production complementarity effects cannot be ruled out, they do not seem to be the main driver of the spillover effects that I find.

Table C.3
Robustness check: dropping non-employment spells.

Dependent variable: $ln(w_{ijt})$	(1)	(2)	(3)	(4)
	Baseline	Drop if ≥ 3 years	Drop if ≥ 2 years	Drop if ≥ 1 year
Experience	0.018***	0.021***	0.021***	0.021***
	(0.000)	(0.000)	(0.000)	(0.000)
Experience^2	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm size / 1000	0.013***	0.015***	0.015***	0.016***
	(0.000)	(0.000)	(0.000)	(0.000)
Coworker Quality $\bar{\theta}$	0.358***	0.314***	0.304***	0.283***
	(0.002)	(0.004)	(0.004)	(0.004)
N	28115529	14441428	13290915	10941628
Mean Log Wages	7.884	7.943	7.952	7.979

Notes: The dependent variable is individual monthly earnings (FTE), in logs. Standard errors clustered at the firm level. Columns (2) (3) and (4) present estimates based on a sample where individuals with longer employment spells are dropped. In column (2), (3) and (4), long non-employment spells are those that last at least three years, at least two and at least one year, respectively. Differences in the effect of 'Coworker Quality' of Columns 2-4 compared to Column 1 are significant at all conventional significance levels. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: Veneto Worker History Dataset, for years 1982-2001.

C.3. Dropping non-employment spells

My baseline regression implicitly assumes that non-employment spells do not play a role in the estimation. In particular, the AKM procedure calculates individual fixed effects using employment spells only, and therefore does not allow for workers' skills to develop differently after long non-employment spells. In this short section, I run the main specification again after having dropped all observations of individuals who have long non-employment spells within the period of my data.

Results are presented in Table C.3. Column (1) reports the results of the baseline regression for comparison. In Columns (2), (3) and (4) I run the main spillover model after having dropped individuals with non-employment spells of significant length and having consequently recalculated our measure of peer quality $\bar{\theta}$. In Column (2), long non-employment spells are defined as those lasting at least three years. I then define long non-employment spells as those that last two years or longer (Column 3) and one year and longer (Column 4). While we lose over 60 percent of the observations between Column (1) and Column (4), the estimates on spillover effects are relatively stable throughout the samples. The tendency of our estimates of spillover effects to fall gradually between column (1) and column (4) is likely to be the result of the fact that the sample is likely to be increasingly positively selected by labour market skills (individuals with more education or skills and higher wages are less likely to have non-employment spells) and that spillover effects may be smaller for individuals with better skills. Indeed, at the bottom of Table C.3 we show that average real log monthly wages increase from 7.88 in Column (1) to 7.98 in Column (4).

References

Abowd, John M., Kramarz, Francis, Margolis, David N., 1999. High wage workers and high wage firms. Econometrica 67 (2), 251–333.

Abowd, John M., Creecy, Robert H., Kramarz, Francis, 2002. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. mimeo.

Altonji, Joseph G., Huang, Ching-I, Taber, Christopher R., 2015. Estimating the cream skimming effect of school choice. J. Polit. Econ. 123 (2), 266–324.

Arcidiacono, Peter, Foster, Gigi, Goodpaster, Natalie, Kinsler, Josh, 2012. Estimating spillovers using panel data, with an application to the classroom. Quant. Econ. 3, 421–470.

Arcidiacono, Peter, Kinsler, Joshua, Price, Joseph, 2017. Productivity Spillovers in Team Production: Evidence from Professional Basketball. J. Labor Econ. 35 (1), 191–225.

Bagger, Jesper, Lentz, Rasmus, 2014. An Empirical Model of Wage Dispersion with Sorting. NBER Working Paper 20031.

Bandiera, Oriana, Barankay, Iwan, Rasul, Imran, 2009. Social connections and incentives in the workplace: evidence from personnel data. Econometrica 77 (4), 1047–1094.

Bartolucci, Cristian, Francesco Devicienti, 2013. Better workers move to better firms: a simple test to identify sorting. IZA Discussion Paper 7601.

Battu, Harminder, Belfield, Clive R., Sloane, Peter J., 2003. Human capital spillovers within the workplace: evidence for Great Britain. Oxf. Bull. Econ. Stat. 65 (5), 575–594.

Becker, Brian E., Huselid, Mark A., 1992. The incentive effects of tournament compensation systems. Adm. Sci. Q. 37 (2), 336–350.

Boning, Brent, Ichniowski, Casey, Shaw, Kathryn, 2007. Opportunity counts: teams and the effectiveness of production incentives. J. Labor Econ. 25 (4), 613–650.

Borjas, George J., 1987. Self-selection and the earnings of immigrants. Am. Econ. Rev. 77 (4), 531–553.

Bramoulle, Yann, Djebbari, Habiba, Fortin, Bernard, 2009. Identification of peer effects through social networks. J. Econ. 150 (1), 41–55.

Brock, William A., Durlauf, Steven N., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68 (2), 235–260.

Card, David, Heining, Jörg, Kline, Patrick, 2013. Workplace heterogeneity and the rise of West German wage inequality. Q. J. Econ. 128 (3), 967–1015.

Card, David, Devicienti, Francesco, Maida, Agata, 2014. Rent-sharing, holdup, and wages: evidence from matched panel data. Rev. Econ. Stud. 81 (1), 84–111.

Chan, Tat Y., Li, Jia, Pierce, Lamar, 2012. Compensation and Peer Effects in Competing Sales Teams. mimeo.

Cingano, Federico, 2003. Returns to specific skills in industrial districts. Labour Econ. 10 (2), 149–164.

Cingano, Federico, Rosolia, Alfonso, 2012. People i know: job search and social networks. J. Labor Econ. 30 (2), 291–332.

Cornelissen, Thomas, Dustmann, Christian, Schönberg, Uta, 2017. Peer Effects in the Workplace. Am. Econ. Rev. 107 (2), 425–456.

Davis, Steve J., Haltiwanger, John, 1991. Wage Dispersion between and within U.S. Manufacturing Plants, 1963-86. Brookings Pap. Econ. Act., 115–180.

Dittrich, Marcus, Knabe, Andreas, Leipold, Kristina, 2011. Spillover Effects of Minimum Wages: Theory and Experimental Evidence. CESifo Working Paper 3576.

Dunne, Timothy, Haltiwanger, John, Foster, Lucia, 2000. Wage and Productivity Dispersion in U.S. Manufacturing: The Role of Computer Investment. NBER Working Paper 7465.

Dustmann, Christian, Glitz, Albrecht, Schönberg, Uta, Brücker, Herbert, 2016. Referral-based job search networks. Rev. Econ. Stud. 83 (2), 514–546.

Edin, Per-Anders, Holmlund, Bertil, Skans, Oskar Nordström, 2007. Wage dispersion between and within plants: Sweden 1985–2000. NBER Working Paper 13021.

Erickson, C.L., Ichino, A., 1994. Wage differentials in Italy: market forces, institutions, and inflation. NBER Working Paper 4922.

Falk, Armin, Ichino, Andrea, 2006. Clean evidence on peer effects. J. Labor Econ. 24 (1), 39–57.

Glitz, Albrecht. Coworker Networks in the Labour Market. Labour Econ., (in press).

Guryan, Jonathan, Kroft, Kory, Notowidigdo, Matthew J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. Am. Econ. J.: Appl. Econ. 1 (4), 34–68.

Hamilton, Barton H., Nickerson, Jack A., Owan, Hideo, 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. J. Polit. Econ. 111 (3), 465–497.

Harbring, Christine, Irlenbusch, Bernd, 2003. An experimental study on tournament design. Labour Econ. 10 (4), 443–464.

Hellerstein, Judith K., Neumark, David, 2008. Workplace segregation in the United States: race, ethnicity, and skill. Rev. Econ. Stat. 90 (3), 459–477.

Ichino, Andrea, Maggi, Giovanni, 2000. Work environment and individual background: explaining regional shirking differentials in a large Italian firm. Q. J. Econ. 115 (3), 1057–1090.

Kramarz, Francis, Lollivier, Stefan, Pele, Louis-Paul, 1996. Wage inequalities and firm-specific compensation policies in France. Ann. D.'Econ. Et. De. Stat. 41-42, 369–386.

Kremer, Michael, 1993. The o-ring theory of economic development. Q. J. Econ. 108 (3), 551–575.

Kremer, Michael, Maskin, Eric, 1996. Wage Inequality and Segregation by Skill. NBER Working Paper 5718.

Kronenberg, Kristin, Kronenberg, Tobias, 2011. Keeping up with the Joneses by finding a better-paid job - The effect of relative income on job mobility. mimeo.

Lazear, Edward P., Rosen, Sherwin, 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89 (5), 841–864.

Lazear, Edward P., Shaw, Kathryn L., 2008. Wage Structure, Raises, and Mobility: An Introduction to International Comparisons of the Structure of Wages within and across Firms. In The Structure of Wages: An International Comparison, Chapter 1, pp. 1–57. Stanford U: A National Bureau of Economic Research Conference Report. Chicago and London: University of Chicago Press.

Lengermann, Paul, 2002. Is it Who You Are, Where You Work, or With Whom You Work? Reassessing the Relationship Between Skill Segregation and Wage Inequality. U.S. Census Bureau, Longitudinal Employer-Household Dynamics Technical Paper 2002–10.

Lopes de Melo, Rafael, 2009. Emparelhamento no Mercado de Trabalho Brasileiro: Evidência e uma Explicação. Estado de uma Nação.

Manski, Charles F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60 (3), p531–542.

Marshall, Alfred, 1890. Principles of Economics. Macmillan, New York.

Mas, Alexandre, Moretti, Enrico, 2009. Peers at work. Am. Econ. Rev. 99 (1), 112–145.

Moretti, Enrico, 2004. Workers' education, spillovers, and productivity: evidence from plant-level production functions. Am. Econ. Rev. 94 (3), 656–690.

Navon, Guy, 2010. Human Capital Spillovers in the Workplace: labor Diversity and Productivity. Isr. Econ. Rev. 8 (1), 69–90.

Occari, Fabio, Pitingaro, Serafino, 1997. Demografia di impresa e mobilita del lavoro:una stima della componente spuria sulla base degli archivi Inps. mimeo.

Ouazad, Amine, 2007. Program for the Estimation of Two-Way Fixed Effects. mimeo.

Piore, Michael, 2009. The Handbook of Industrial Districts. In: Becattini, G., Bellandi, M., Propis, L.D. (Eds.), Conceptualizing the Dynamics of Industrial Districts. Edward Elgar, Cheltenham, UK; Northampton, MA, USA.

Shvydko, Tetyana, 2007. Interactions at the Workplace: Peer Effects in Earnings. mimeo.

Tattara, Giuseppe, Valentini, Marco, 2010. Turnover and excess worker reallocation: the Veneto labour market between 1982-1996. Labour 24 (4), 474–500.

Wooldridge, Jeffrey M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.